Covid-19 Mortality Rates and Housing Prices

Adam Finch

12 April 2022

Econometrics (ECON 388)

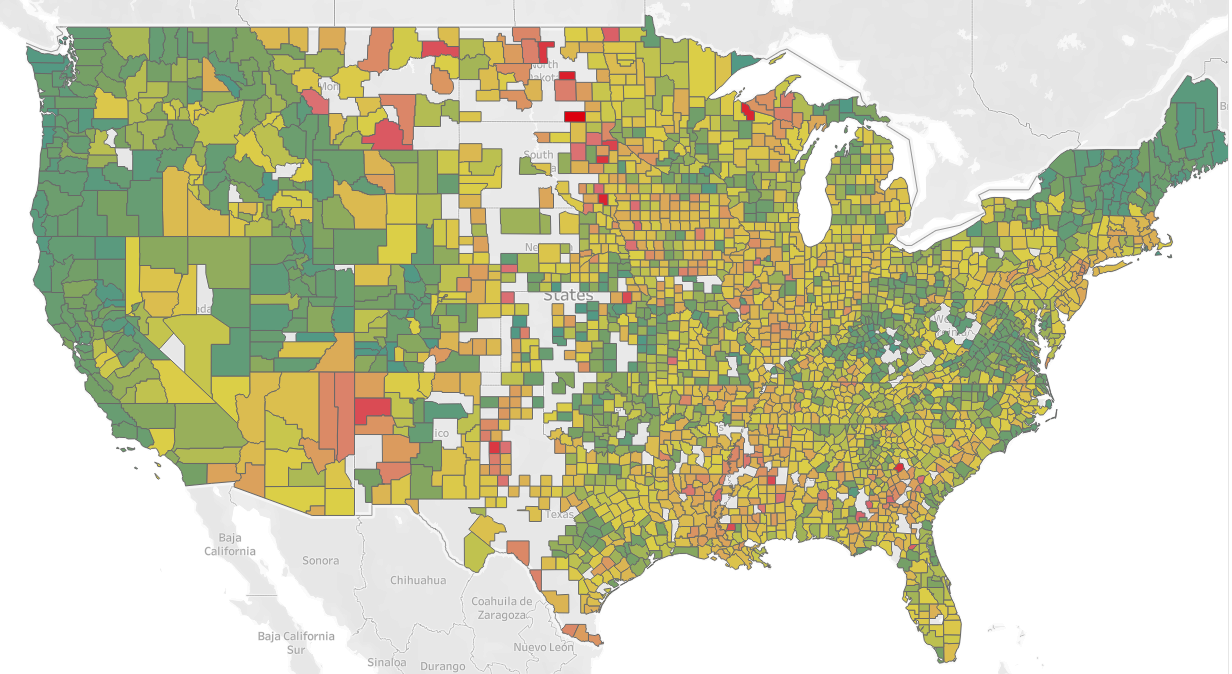
***Datasets, my STATA do file, and other resources can be found at:* https://github.com/afinch99/ECON388DA3**

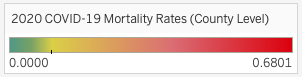
**Introduction**

Covid-19 presented the largest public health crisis in recent history. Millions of lives were lost, unemployment soared, and financial markets crashed. Recently, declining case counts have produced optimism that the nation is nearing the end of this catastrophe. However, in some ways this recovery period may be just as difficult as the pandemic itself. One specific area of concern is rising housing prices across the United States. There are many possible causes for these increased prices. This memo seeks to evaluate one of these possible causes by answering the following research question: what effect did county-level per-capita Covid-19 mortality rates in 2020 have on housing prices in 2021? There are several potential hypotheses that attempt to explain the relationship between these two variables. Perhaps housing prices decline in areas that have a high mortality rate because prospective home buyers are deterred from moving to areas that have not implemented precautionary measures such as mask mandates to slow the spread. On the other hand, it is also possible that housing prices increase in areas with a high mortality rate because potential home buyers are attracted to communities with relaxed Covid-19 related regulations. Analyzing census, Covid-19, and housing price data will help to address these hypotheses and identify whether or not a strong relationship exists between Covid-19 mortality rates and housing prices at the county level.

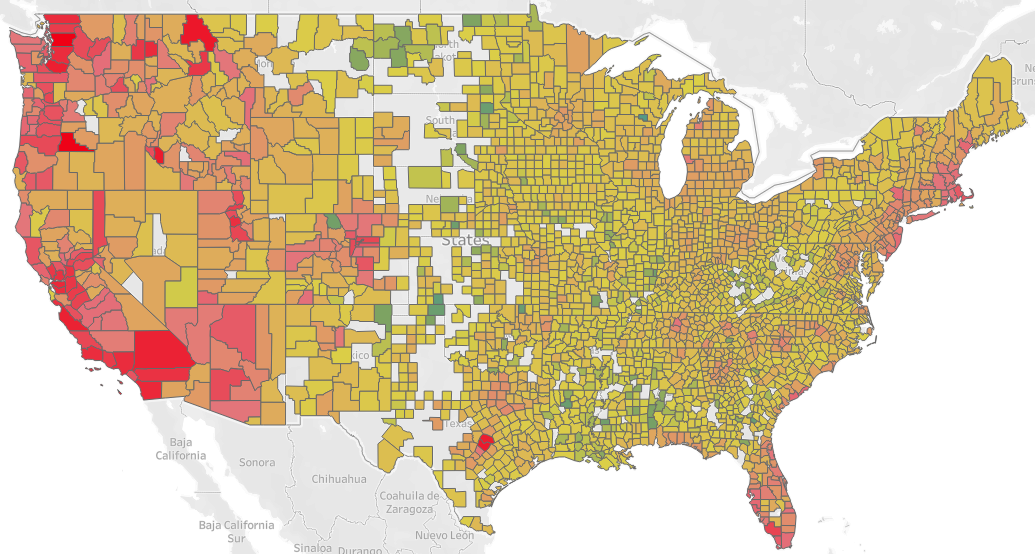
**Data**

In order to answer the central research question, I needed to collect data from several different sources. The first dataset contained information regarding 2020 Covid-19 cases and deaths in each U.S. county. It was obtained from the New York Times Covid-19 database. The second dataset came from the 2020 U.S. Census; the most useful information in this dataset was 2020 population estimates at the county level. Merging the New York Times dataset with the U.S. Census dataset made it possible to generate the 2020 Covid-19 mortality rate at the county level. Finally, the third dataset contained 2020 and 2021 Housing Price Index (HPI) numbers for counties in the United States. This data is publicly available on the Federal Housing Finance Agency’s website.

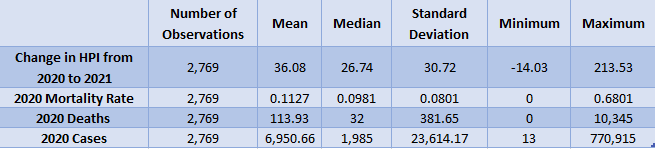
These three datasets included many different variables. One variable of interest that I chose to analyze is mortality rate. I generated this variable by aggregating the daily Covid deaths in each county for the year 2020 and dividing the total deaths by the total population, multiplying the result by 100. The following map -- generated in Adobe Tableau -- is a visual representation of the different mortality rate levels among U.S. counties included in the final dataset. As the key indicates, counties shaded in green had comparatively low levels of Covid-19 related deaths per capita in 2020, and counties shaded in orange and red experienced higher levels. Areas on the map that were left unshaded are counties that were not included in the final dataset. 



I also generated a similar map demonstrating the change in HPI from 2020 to 2021; it is shown below. This map shows that there were very few counties where housing prices declined from 2020 to 2021. One interesting take away from comparing these two visualizations is that in general, areas with noticeably low Covid-19 Mortality rates compared to the rest of the country seem to have experienced higher increases in HPI from 2020 to 2021.

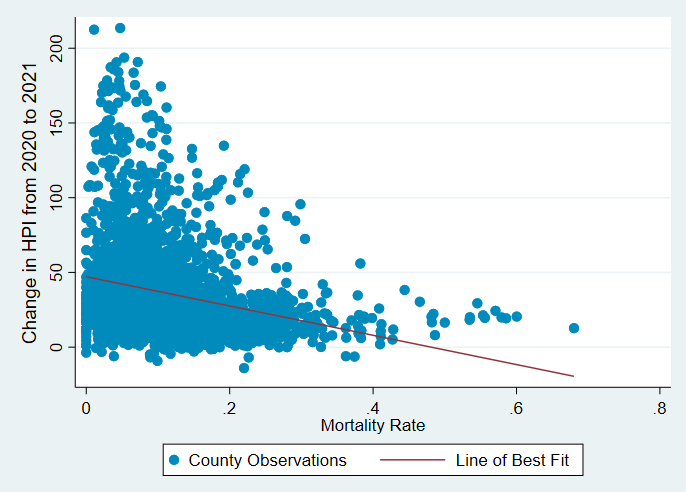


Finally, the following table represents a summary of several interesting variables. This table also demonstrates why it was important to use mortality rate instead of total deaths or cases: very highly populated counties can skew the average for a variable. This is shown in the large difference between the mean and the median for the variables deaths and cases. Because there is such a wide range of population levels between U.S. counties, it was necessary to generate the standardized per capita metric of mortality rate in order to make comparisons.



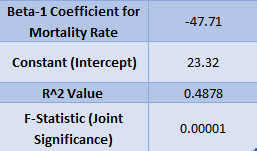
**Empirical Analysis**

In order to determine the effect that Covid-19 mortality rate had on the change in HPI from 2020 to 2021, it was necessary to run statistical regression tests. A simple regression of changes in HPI on mortality rate resulted in the following regression equation and scatterplot:

*PredictedHPIchange =* 47.12 - 97.96(*MortalityRate)* 

This regression indicates that a county with a mortality rate of 0 would have a predicted increase in HPI of 47.12 from 2020 to 2021. As the mortality rate goes up, the model predicts that HPI will decrease at a slope of -97.96. For example, suppose a county had a mortality rate of 0.2, meaning that 0.2% of the county’s population died from Covid-19 in 2020. This simple regression model would predict the change in HPI for that county from 2020 to 2021 to be 27.53.

A p-value of 0.0001 indicates that these results are statistically significant. However, it is important to note that correlation does not necessarily mean that causation is present. Additionally, an R-squared value of 0.0653 shows that knowing a county’s mortality rate explains only 6.53% of the variation in HPI change. In order to improve this number, and also to more plausibly show causation, I ran another regression. This time I generated a dummy variable for each state in order to control for the effects that a county being located in a particular state might have on HPI change from 2020 to 2021. Controlling for state was important because for some counties, simply being associated with a certain state might heavily influence the HPI change.

As predicted, the results of this regression showed that what state a county is located in heavily influences the predicted HPI change. The regression results are shown below: 

The first takeaway from these results is that this model explains much more of the variation in HPI change than the first model did. This is demonstrated by the R-squared value, which implies that knowing a county’s mortality rate and what state they are in explains 48.78% of the variation in HPI change from 2020 to 2021. Second, the coefficient of -47.71 is much more likely to accurately explain the specific effect that mortality rate has on HPI change than the coefficient of -97.96 from the first model. This model is also more likely to satisfy the assumptions necessary for a regression to have a causal interpretation. The assumption most likely to be violated is the zero conditional mean assumption, which states that the right-hand side variables cannot be correlated with the error term. Although controlling for states took out one potential correlation between the right-hand side variable of mortality rate and the error term, there are other factors included in the error term that are likely to be correlated with mortality rate. Examples of things that should probably be controlled for are the age of the population in a county, the average income within a county, and the healthcare options within a county. Including these variables will help to correct for potential bias within the regression model.

**Conclusion:**

Overall, I observed that there is a negative correlation between mortality rate and HPI change from 2020 to 2021. On average, a one percent increase in mortality rate is correlated with a 47.41 unit decrease in HPI when controlling for the state that a county is located in. This correlation is statistically significant and explains a fair amount of the variation in HPI change among counties. A causal relationship is certainly plausible, although to fully satisfy the zero conditional mean assumption more variables should be taken into account. This is one limitation of the data set; it does not include relevant variables such as age, income, and healthcare structure. Inflation is another important factor that certainly has an impact on housing prices. Further analysis should involve collecting county level data on these variables and adding them to the regression model. As the United States exits the pandemic, analyzing Covid related data is important. Identifying key trends can help explain the past, remedy present problems, and prepare the nation for future public health emergencies.